COMMEMORATIVE/EVENT REPORT

JULY 2026

DRUMS 2025

Information Manipulation and Interference Against Technological Advancements, Shifting Global Power Dynamics and Ongoing Conflicts

RSiS S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES
Nanyang Technological University, Singapore

RSiS 30

Contents

Report on the workshop organised by: Centre of Excellence for National Security (CENS) S. Rajaratnam School of International Studies (RSIS) Nanyang Technological University, Singapore

Rapporteur: Yasmine Wong, Tan E Guang Eugene, Davis Zheng, Sean Tan, Juan Cui Ying and Andrea Luz "Oona" Nery

Editors: Gulizar Haciyakupoglu and Dymples Leong

Executive Summary

The Centre of Excellence for National Security (CENS) held its annual Disinformation, Rumours, Untruths, Misinformation and Smears (DRUMS) conference from 28 to 29 October 2025.

The conference theme was "Information Manipulation and Interference Against Technological Advancements, Shifting Global Power Dynamics and Ongoing Conflicts". Over two days, 230 participants from government agencies, academia, diplomatic corps, and non-governmental organisations learned from and engaged with 13 international and local speakers.

The keynote speech set the tone for the conference by exploring how foreign malign influence is entering its second decade with the deployment of generative Artificial Intelligence (AI) and the potential impact of a new era of AI-first actors, who have built their processes around generative AI tools to create content for corporations, political groups, nation-states and others.

Panel 1 speakers spoke on trends in the information landscape. They touched on topics including the public perception of online falsehoods and AI, as well as cases of political propaganda and information manipulation across Brazil, India, and Europe.

Panel 2 speakers spotlighted the usage of AI in strategic communications, identifying various opportunities and vulnerabilities AI can bring, especially with the rapid development of generative AI.

Speakers for Panel 3 shared global case studies on information manipulation and interference, and discussed new strategies, including grey zone tactics and weaponisation of emerging technologies.

Panel 4 speakers surfaced valuable strategies and insights to manage disinformation. They stressed how building trust and resilience against disinformation can be conducted through collaboration, transparency, and adaptive governance in an evolving digital landscape.

The 2025 edition of DRUMS received positive feedback on various aspects, including the design and organisation of panels, selection of speakers and topics, operation, and administrative matters.

The following sections of this report summarise key points from the speakers' presentations. Key takeaways from the Q&A sessions are available at the end of each panel.

Keynote Speech

The Second Decade of Malign Influence

Clint Watts, *General Manager, Microsoft Threat Analysis Center*

- In the early days of the internet, technology largely served to connect people. The current digital era has seen the threat landscape evolve from individual hackers and small-scale cyber criminals to large, coordinated attacks carried out by nation-states and organised cybercrime groups.

- Between 2005 and 2006, as the world focused on the aftermath of the 9/11 terrorist attacks and the Global War on Terrorism, Al-Qaeda in Iraq began surfacing on YouTube. Within just five years, this isolated digital presence evolved into a standard, as nearly every extremist group has gone online and on social media to conduct their operations.

- While the Arab Spring had initially renewed hope for technology, the optimism faded shortly thereafter. This disillusionment was partly stemmed from a failure to distinguish between social media actors and cybersecurity actors. While the cybersecurity landscape faces Advanced Persistent Threats (APTs) - a long term targeted cyberattack, designed for infiltration of organisations and to remain undetected for as long as possible - social media platforms face a parallel change with Advanced Persistent Manipulators (APMs), which are well-resourced and determined entities with scaled operations that may span the entire spectrum of online and offline operations.

- While social media has facilitated connection, it simultaneously created divisions in society and changed how individuals interact with one another. Algorithms of social media that prioritise personal engagement have driven this change inadvertently fuelling confirmation bias.

- Russia, China and Iran's influence operations have been focusing on the following:

    - Russian threat actors have shifted their attention towards Armenia and Moldova in recent months, after previously concentrating on the US and UK, and the 2024 Paris Summer Olympics. The Russian propagandist group Storm-1516 was involved in the Neva Flood cyber influence operations, which centred on the 2024 Harris-Walz campaign in the US and utilised face swapping. Storm-1516 often uses fabricated whistleblower narratives to criticise and discredit a specific country. Storm-1516 took a considerable amount of time to create videos, despite their knowledge and understanding of its intended target audience. While this slower production timeline was common in the past, the scale of content production and volume of online information could lead to of the risk of content being overshadowed before it gains traction.

    - Information manipulation efforts by China and Russia are different. Whilst China's disinformation efforts get a lot of coverage in the media, they have not been achieving tremendous impact. China's state media, which is available across the globe, is more impactful.

    - Iran has not been very successful with propaganda and disinformation. Their focus has been on regional conflicts, occasionally focusing on elections (albeit poorly executed). The war with Israel has dominated their efforts in the past few months. Iran utilises any AI tools available to them.

- There are now AI-first actors, individuals or groups that design influence operations around generative AI tools, as the primary means to create content for corporations, political groups, nation-states and others. AI-first actors are well-resourced and have the technical know-how to use the tools well. While AI may facilitate the creation of large volumes of propaganda, the quality of the content remains questionable (i.e., AI slop). Low quality may gradually reduce the engagement as engagement with singular pieces of material seem to be lesser.

- Although AI adoption has increased, it is difficult to use AI convincingly, especially with quick identification of its inauthenticity. Audiences tend to gravitate towards authentic content. Audio AI has been more impactful for propaganda and disinformation, relying on fewer training datasets and less computational power to generate content. Factors influencing the credibility of the content include:

    - Familiarity with the impersonated person: If a familiar person is being impersonated, people tend not to fall for the deepfake.
    - Timing of misinformation circulation: During times of calamity, misleading and conflicting information could be rife.
    - Whether the information is consumed in a private setting: If the information is consumed in a private setting with no one to verify the information with, the person may be more likely to believe in the deepfake.

- Access can shape the threat. Access points (the channels, technologies or platforms through which people can encounter online harms), .and the face of threats have changed with mobile phones, which increased the avenues for radicalisation. When looking at AI threats, it is important to note that most people still do not have access, or face limitations in technology use (e.g., powerful internet connections, bandwidth to utilise tools). In the first era of the Internet, cyber criminals were the first to utilise AI, with individuals then scaling up to operationalise collectively. As access grows and data centres continue to be built across the world, the nature of threat actors evolve beyond individual cybercriminals, to threat teams at the level of corporations with inter-platform collaboration.

- Individuals are increasingly able to use AI to communicate with others, and to curate and shape their own information environments in ways that reinforce their confirmation and implicit biases. Audiences are both savvier and more informed on AI-related threats, with increasing discussions surrounding the implications of AI. With the overwhelming amount of content and little way to filter and handle the information overload, younger generations may see a move back to analogue.

Key points noted from the Q&A session

Issue 1: The nature of influence operations has changed.

Convincing and successful influence operations require substantial resources. Russia has spent a large amount of money to maintain its influence operations worldwide. Another important factor is understanding the audience. In the past, threat actors used to mine social media to identify patterns of vulnerabilities and target specific groups. This approach is limited as it eventually shifted towards recruiting people within the target audience to serve as influencers.

A threat actor's connection with the target audience, such as shared language or political alignment, enhances their ability to reach a broad audience and increase their impact. Consequently, the existence of a sizeable diaspora helps expand the reach of their influence.

Issue 2: The information landscape can potentially be poisoned by flooding it with information.

Ten years ago, bots were used to amplify existing user-generated content. Now, bots themselves can put out content and respond. In this context, AI is used to simulate a variety of actors and create engagement with content. Differing rules in each country pertains to the boundaries around propaganda and disinformation (i.e., what is acceptable behaviour and what is not).

Issue 3: Influence of user engagement on algorithms.

Current algorithms continue to rely primarily on user engagement metrics (e.g., likes) to suggest similar materials. Audience preference takes precedence over algorithmic manipulation and actors engaging in information manipulation have to consider this to reach audiences.

Panel 1: Trends in Information Landscape

Restoring Information Integrity: Actors, Action and Accountability

Dr Carol Soon, *Associate Professor, Department of Communications and New Media, National University of Singapore*

There have been structural shifts in the media landscape that have made information integrity more fragile, including the professionalisation of influence. The media landscape has become increasingly fragmented with traditional media, social media and a growing number of independent content creators competing for public attention. Various organisations, including the Organisation for Economic Co-operation and Development (OECD), have endeavoured to restore information integrity, aiming to ensure access to accurate, reliable, evidence-based, and diverse information sources.

- *Political actors and disinformation during election campaigns:* Political actors increasingly rely on social media consultants, public relations and advertising strategists to boost their profile during election campaigns. Some of these campaigns involve creating fake Facebook pages and accounts circulating misleading and harmful information, including scandalous information on opponents.

- *Risks concerning the use of AI during election periods:* The growing use of AI during election periods could amplify the illusory effect of political statements. The use of AI in a seemingly non-malicious manner to soften a candidate's image—such as portraying them as a benign, friendly character—could distract the electorate from substantive policy and "bread-and-butter" discussions.

- In the context of the use of AI in the run up to the 2025 General Elections in Singapore, a survey of data revealed that one in three respondents encountered AI-manipulated content regularly. There are three possible reasons why the exposure was not higher:

    - Restraint: Actors did not abuse the technology to a significant degree.
    - Regulation: The Elections (Integrity of Online Advertising) (Amendment) Act 2024 acted as a legal deterrent.
    - Sophistication: Some content is realistic to the degree that some voters could not distinguish it from reality.

A study on digital trust suggests that the digital trust of technology involve sevens dimensions: cybersecurity, safety, transparency and accountability, reliability, fairness, redressability, and privacy. Trust in the organisation behind the product is also a fundamental component of the digital trust. Part of this trust concerns the mechanical aspects of technology, including its reliability. The study also revealed that the trust in the protection of privacy and transparency in using the technology is lower when compared to other metrics. On the rollout of technology products, the trust in the Singapore government is the highest, followed by technology providers and non-government organisations.

Artificial Intelligence and Extreme Speech: Is the Hype Justified?

Dr Sahana Udupa, *Professor of Media Anthropology, Ludwig-Maximiians Universitat, Munich*

AI is seen as a transformative force in misinformation dissemination, enabling deepfakes, predictive analytics and micro-targeting. However, the AI hype clouds the complexities concerning how extreme speech operates. Information dissemination and amplification mechanisms have a greater impact on extreme speech than the sophistication of AI technologies, and the impact of deepfakes, including during elections, is questionable.

- Deepfakes challenge the referential link (the connection between audio-visual content and the real-world events) between audio-visuals and reality - what we see/hear and what is real, impairing the truth and evidence paradigms. Developments in agentic AI (autonomous systems) could exacerbate this disruption, leading to the emergence of more contaminated audio-visual content. Some of the current AI fakes are socially shallow and can be easily debunked. However, they gain traction when they tap into powerful underlying narratives that resonate with revivalist ideas among people or a nation.

- Deepfakes can be Thin or Thick Fakes:

    - Thin Fakes can use humour and may contribute to the repertoire of internet memes and witty campaign styles.
    - Thick Fakes may involve AI manipulations, low technology cutting and mixing, real-world captures like those taken from a phone, or a combination of these.

- The distribution of disinformation may involve professional political consultants or disinformation services that offer data-based solutions to reach political groups that were hitherto unreceptive to the candidates, through tailor-made political strategies or messages. These dynamics may also extend to non-official influencers that may support a particular campaign that are not officially sanctioned but are elaborately financed. These groups utilise WhatsApp extensively, especially for intrusive propaganda where party people organically embed themselves within trusted WhatsApp groups as part of their community reach.

- This distribution pattern of embedded influencers in chat groups allows the tapping on community trust, where people may feel socially obligated to respond and affirm potentially false and extreme narratives. Additionally, members of the group may also feel obligated to share these ideas as responsible members without a critical analysis or correction of the message, especially if it is a message coming from a senior member of the family.

Diplomacy Squared: Analysing the Kremlin's manipulation tactics in physical and digital international fora

Ben Schultz, *Lead Researcher, American Sunlight Project*

Russian information manipulation efforts in the digital and physical arena have intensified since Russia's invasion of Ukraine. New Foreign Information Manipulation and Interference (FIMI) techniques have emerged as Russian bot operations continue. New techniques include Large Language Model (LLM) grooming, co-option of international forums and the evolution of bot networks.

- On platforms like X, Russia utilises "sleeper agent" bot networks that remain dormant until they are mobilised to conduct influence operations. These bots may be called upon to engage with and repost Kremlin's content. Some bot networks may share content on non-contentious issues, such as environmental protection, to impede their identification as conduits of influence. Russian botnet accounts target specific institutions and their interactions in Europe and the US.

- The Russian botnets engaging in FIMI can be identified through their rapid reposting of statements from Russian diplomatic accounts. They can also be detected through Russian words embedded in messages and codes, and the coincidental timing of posts with Russian internet outages. Some botnet accounts have posted over a million times on X, demonstrating a high level of reach for a brute-force bot network.

- LLM grooming can help shift the narrative in Russia's favour. LLMs like ChatGPT, Gemini and DeepSeek pick up articles published in less common languages around the world, such as Gaelic or Māori, by Kremlin-controlled websites, and offer them as legitimate search results. This network has been expanding from Europe to other countries such as Singapore.

- Russia is also pursuing international lawfare. For example, Russia alleged that biolabs in Ukraine were undertaking dangerous, experimental research on mosquitoes, bats and birds. As a result of this accusation, the topic was placed on the United Nations' agenda for discussion. A law firm based in Moscow has filed thousands of cases before the European Court of Human Rights, advancing false narratives to manipulate the court's case system and waste court resources. The court has recently rejected the first of these cases.

- Other disinformation networks also engage in similar activities. The Kremlin-run Matryoshka network replies to academics and journalists, asking them to fact-check a certain issue, wasting their time. Journalists have been sued for factual reporting, which encroaches on freedom of expression and academic freedom.

Key points noted from the Q&A session

Issue: Technical and human friction can help slow the spread of misinformation, while disinformation playbook can be leveraged to share research.

Friction that social media platforms could introduce through technical means includes labelling content and limiting the number of times a message can be forwarded. However, such measures have had only a limited effect on sharing. Human friction occurs when members of the chat group react against forwarded messages and stop sharing them – they may also ignore the sender. In addition to friction, the playbook against disinformation can be used to spread the research and narratives through active sharing by interested parties and reposting to maximise impact.

Issue: Disinformation campaigns have become less resource intensive and disinformation-for-hire services are accessible to various actors.

Disinformation campaigns have become less resource-intensive, with the availability of overseas troll farms that have the human capacity to run them. However, threat actors may not have a steady stream of revenue to sustain these operations. Influencers on social media can easily buy large numbers of followers to amplify their reach. Russia is not the only country that has used disinformation-for-hire services; other authoritarian states have leveraged them as part of their information operations as well.

Issue: Actors and their Tactics, Techniques and Procedures (TTPs), as well as distribution techniques need to be dealt with as a first instance when countering DRUMS.

Content moderation is a difficult task, and it is hard to pinpoint which terms should be moderated. It is easier to target distribution networks and actors that perpetuate the disinformation. These include social media consultants, public relations advertisers, and political marketers. Instant messaging apps can also be pressed upon to debunk misinformation and slow down its spread in closed groups.

Panel 2: Use of Artificial Intelligence for Strategic Communications

AI Challenges, Immediate Responses: Building Societal Resilience Through Strategic Communication

Darius Buta, *Senior Advisor - National Crisis Management Centre of the Government of the Republic of Lithuania*

Lithuania's National Crisis Management Centre (NCMC) plays a critical role in crisis management and managing disinformation threats.

- National security threats include airspace violations, terrorist attacks and arson, which mostly originate from Belarus, Russia and China. The NCMC is responsible for monitoring threats, coordinating national security, and organising exercises such as the 2025 NATO Crisis Management Exercise.

- Lithuania's operational crisis management strategy model comprises a cross-institutional team of analysts and strategic communication experts from various organisations. NCMC aggregates data from various government agencies and organisations to coordinate information sharing and inform key decision-makers during crises. This includes establishing real-time communication with media organisations to help debunk disinformation.

- "Kaliningrad transit" is a disinformation case that involved coordinated inauthentic behaviour by bot networks, spreading the fear that Lithuania would deny logistics transfers from Russia to Kaliningrad. Facebook took down a portion of the false information after significant effort by NCMC.

- Countering disinformation is not a singular organisational effort. It requires everyone to work together. AI is part of the future, but trust is not built by machines or AI; it is ultimately built by people.

How AI Shapes and Warps Strategic Communication

Dr Surachanee "Hammerli" Sriyai, *Visiting Fellow - ISEAS, Yusof Ishak Institute*

During the 2025 Thailand and Cambodia border dispute, both countries engaged in strategic communication, including narrative framing. They both sought to capture the narrative during the period of conflict by presenting their own views of the situation.

- The international humanitarian law has clear guidelines concerning state behaviour. However, it does not provide any guidelines for how states should treat non-state actors - states may choose to persecute them, ignore them or consider other actions. International humanitarian law also does not expressly prohibit information influence operations, including disinformation, during armed conflicts unless they constitute perfidy, incite war crimes or directly target protected persons or objects. Strategic communication is generally used in a state-centric approach.

- Thailand's strategic communication efforts were based on traditional media and official spokespersons, and there was a lack of coordination between the defence, digital and foreign affairs ministries regarding information control. Official announcements lagged viral narratives trending on social media, which were shaping regional perception. Non-state actors filled communication voids traditionally occupied by the state, spreading false narratives during lull periods where no public announcements were made. Cambodia was the first to officially release a statement. Cambodia was perceived to be non-transparent regarding the death counts of its soldiers in the conflict.

- While neither country leveraged AI, non-state actors employed it to great effect. Among others, non-state actors used AI to discredit key figures during the crisis, including a deepfake image of Cambodia's spokesperson, Maly Socheata, being overtly friendly to Hun Sen, Cambodia's current president of the senate. There was no integrated AI system for real-time narrative tracking. The use of AI could have provided early warnings of a nationalist surge or been used for counter-messaging and disinformation monitoring. Netizen involvement in the conflict and fact-checking efforts have led to information and counterinformation spreading much faster than state-led efforts.

- There is a gap in AI governance. Current international law, as applied to cyberwarfare, predicates state responsibility on clear attribution. In cases of operations by non-state actors, this has faltered and not been effective. Thailand's current approach to AI governance was influenced by the European Union's AI strategy and governance. However, it lacks the same depth and enforceability concerning generative AI, with little to no mention of platform responsibility.

How The Straits Times Uses AI and How It Doesn't

Jeremy Au Yong, *Associate Editor, The Straits Times*

The Straits Times (ST)'s integration of AI into the editorial process is set upon clear parameters of use. Policies pertaining to AI usage include ST's living policy guidebook for regulating AI use and ST's AI hedge, which involves a three-step implementation framework.

- The AI boom saw substantial market investment driving rapid adoption while simultaneously prompting widespread warnings of a potential market bubble. The prevailing fear of missing out surrounding generative AI had compelled businesses to commit vast financial resources to it.

- ST acknowledges that regulation and safety must come before technology. ST believes it is important to regulate AI to prevent misuse even before it is required to do so by government regulations. The organisation has spent a significant amount of time making a living policy playbook to regulate internal AI use, while allowing teams space to experiment.

- ST has come up with the concept of the AI hedge, which is a 3-step implementation framework that defines what the core of ST's no-AI policy involves. ST has a document that clearly outlines which activities, processes and content will remain 100% human-only. ST does not use AI for factchecking. The outsourcing of factchecking could erode the key values of ST. ST has also been emphasising the human core by organising in-person meetings with students and the community to build trust and explain what journalism entails.

- ST has developed a few tools over the past few years. One of the public-facing tools including an AI summary tool and a career chatbot. There are also various internal tools such as the SEO (Search Engine Optimisation) scorecard, which helps to review and increase the search engine optimisation of the articles.

Key points noted from the Q&A session

Issue: Lithuanian government seeks to provide a unified stance against information attacks and collaboration is central to its approach.

Lithuania's Government Resolution No. 955 is a 10-point scale to evaluate serious information attacks. Low-scoring incidents (2–3 points) may not receive a response, as reacting could unintentionally amplify the opponent's message. The government provides a common stance that all organisations echo to maintain a unified front. A recent example was the Vilnius plane crash incident, where the Pravda network (a pro-Russian influence operations network) falsely claimed the aircraft carried ammunition for Ukraine, requiring a coordinated debunking effort. A major emphasis of the framework is building a strong multi-institutional team, with expertise from intelligence, the armed forces, and other relevant fields. This is essential for the NCMC's monitoring and counter-disinformation duties. Additionally, to encourage cohesive cooperation among non-government organisations and civil society, Lithuania focuses on restoring trust and ensuring transparency. Against the legacy of Soviet-era distrust and the importance of not assigning blame when issues arise, NCMC was established to coordinate efforts constructively. The goal is to counter disinformation and strengthen overall societal resilience.

Issue: During the Thai–Cambodia conflict, both countries followed a similar propaganda playbook, aiming to speak first to shape the narrative, and social media only amplified these existing tactics without improving their responsiveness.

Communication on the conflict mainly originated from official state representatives. Both sides used the same long-standing strategy of being first to frame the situation, knowing that only a few people at the front line understood what was really happening. Cambodia pushed its narrative early, even if untrue, and reinforced it repeatedly. Social media merely amplified the state-backed narratives, rather than making their operations more responsive or adaptive.

Issue: The Straits Times places humans at the centre of reporting and it seeks to personalise the news according to its target audience.

ST plans to maintain a strong human element in its reporting, recognising research indicating audience distrust in AI-written news. So far, the organisation has not faced any claims that its published stories were AI-generated. However, it has internally identified and intercepted AI-written pieces before publication. In today's information overload, newsrooms rely on audience targeting and personalisation to attract interest. ST is still working to refine this approach. They produce a wide range of stories aimed at different audience segments, trying to match content to individual interests. However, successful personalisation demands extensive data, making the process difficult to execute effectively.

Issue: Although AI generated Lithuanian content is clearer than a year ago, non-English-speaking countries should consider the potential biases and security risks that could be embedded in the AI tools developed in Western contexts when adopting them.

Recent improvements now allow AI to produce accurate and clear speeches in Lithuanian. Countries purchasing AI tools from Western vendors must be careful as these systems may carry inherent biases. Non-English-speaking states need to be deliberate and cautious in how they integrate AI technologies.

Issue: There are challenges to regulating social media platforms and holding them accountable.

Despite attempts to introduce stricter measures, such as Lithuania's proposal to criminalise bot activity, EU-level resistance limits how far governments can go. While countries can push for accountability, the EU is uniquely positioned to draft and enforce legislation as it represents the regional bloc and can collectively pressure platforms. Lithuania's effort to criminalise social media bot activity was rejected by the EU Commission, showing how difficult it is for individual states to enact stricter controls within regional bloc groupings. Social media companies have grown into major political actors. Facebook's user base exceeds the population of many nations. People may defend access to platforms like TikTok even when states consider bans. AI labelling practices have weakened, and major platforms like Facebook have shifted responsibility to users instead of maintaining safeguards themselves.

Panel 3: Global Case Studies on Information Manipulation and Interference

Information Manipulation and Interference in the South Pacific: Lessons from New Caledonia and Vanuatu

Dr Jean-Baptiste Jeangène Vilmer, *Ambassador of France to Vanuatu and the Solomon Islands*

While geopolitical rivalries, including activities related to FIMI, are often perceived as the exclusive domain of major powers, their repercussions extend significantly beyond these primary actors. The impact is evident in peripheral regions such as the South Pacific, which encompasses several current and former French overseas territories.

- Malign actors are incentivised to conduct information manipulation within the South Pacific for several strategic reasons:
    - *Asymmetric Vulnerabilities:* Small states and overseas territories often have fragile media ecosystems and constrained investigative capacities, rendering them susceptible to external interference.
    - *Strategic Reach:* France's expansive global presence provides adversaries with opportunities to exploit distant vulnerabilities, and challenge French sovereign interests globally.

- Recent history illustrates how bilateral tensions between sovereign states can manifest as influence operations in third-party territories. Notable examples concerning South Pacific include the following:
    - In response to France's diplomatic stance on the Nagorno-Karabakh conflict, Azerbaijan launched influence operations targeting New Caledonia. The Baku Initiative Group (BIG) was identified as a primary instrument in these efforts.
    - Influence operations targeting Vanuatu were part of a broader dissemination of anti-French rhetoric originating in Burkina Faso. These campaigns lionised the Burkinabe leader, Ibrahim Traoré, as a symbol of anti-colonialism. Key actors included figures such as Sissoko Sora Damba, who was highly active on the X platform.

- Information operations in the South Pacific typically involve a multi-pronged approach. Below are some examples utilising different measures:
    - The BIG has actively funded travel expenditures for independence activists and provided academic scholarships. The Azerbaijani leadership openly acknowledged these operations, framing them as gestures of anti-colonial solidarity.
    - In May 2024, during a period of violent civil unrest in New Caledonia, digital monitors observed a surge in online activity. Within a 48-hour window, thousands of coordinated posts appeared across TikTok, X, and

Facebook. Concurrently, content from Benin activist Kemi Seba containing anti-French sentiment was amplified to West Africa.
    - The BIG has disseminated AI-generated video and photo montages purportedly depicting French police brutality, and were amplified through strategic hashtags, such as #RecognizeNewCaledonia and #EndFrenchColonialism, to maximise international reach.
    - These operations relied on consistent themes, primarily portraying France as a colonial oppressor that continues to exploit regional resources. These narratives were often boosted by allegations of systemic racism and the persistence of colonial-era policies.

- Operational patterns identified in Burkina Faso-linked campaigns include:

    - The BIG has disseminated AI-generated video and photo montages purportedly depicting The deployment of bot networks and sock-puppet accounts for coordinated amplification.
    - Strategic partnerships with Pan-African influencers and media outlets to increase visibility.
    - The use of Virtual Private Networks (VPNs) to circumvent content moderation and geo-blocking protocols.

- Countermeasures against information operations should consider the following:

    - It is essential to maintain a balanced perspective, ensuring that the scale of foreign influence is neither understated nor exaggerated.
    - Timely intervention is critical. The swift publication of fact sheets and the suspension of malicious accounts, such as those facilitated by watchdogs such as VIGINUM (the French government agency tasked to detect and analyse foreign interference in France), are vital to neutralising disinformation.
    - Investment must be directed towards bolstering investigative journalism and supporting civil society in vulnerable regions.
    - Developing robust, multilingual communication strategies is necessary to effectively counter adversarial narratives on a global stage.

TikTok and the Disruption of Political Troll Farms and Cybertrooping in Malaysia

Dr Benjamin YH Loh, *Lecturer, Monash University Malaysia*

Malaysia has contended with organised cyber trooper activity for over two decades. This has fundamentally shaped the nation's digital political discourse and influenced several electoral outcomes. The information manipulation measures observed in Malaysia during the early 2000s are largely viewed as precursors to the sophisticated manipulation strategies that are prevalent globally today.

- The cyber troopers are generally classified into three distinct categories based on their level of engagement and motivation:

- *Low-engagement cyber troopers:* These entities focus on the foundational tasks of content amplification, primarily through the systematic reposting of established party narratives to increase visibility.
- *High-engagement cyber troopers:* These sophisticated online personas engage in proactive and often aggressive debate, frequently utilising tactics to destabilise and discredit political opponents.
- *Ideologically motivated actors:* Often referred to as "political die-hards," independent individuals utilise tactics identical to professional cyber troopers, but operate without financial remuneration and are driven instead by personal conviction.

- Prior to general election of 2018, a significant portion of cyber troop operations were conducted by salaried government employees. The landscape has since shifted due to fiscal constraints and political volatility. Under the administration of Mahathir Mohamad, extensive civil service retrenchment led to a marked decline in organised state-sponsored digital operations.

- In the contemporary environment, political influencers have re-emerged with renewed vigour, shifting their focus towards high-impact platforms such as podcasts, X (formerly Twitter), and TikTok. TikTok, in particular, served as a primary catalyst for political disruption during the 2022 general election, where the Parti Islam Se-Malaysia (PAS) utilised the platform to mobilise rural constituencies. The unique algorithmic structure of TikTok necessitates authentic and direct engagement, rendering traditional bot-driven strategies increasingly obsolete in favour of visible, human-centric influence.

- The narratives currently dominating Malaysian disinformation and influence efforts involve the following themes:

  - *Ethnicity and Religion:* These remain the primary pillars of political mobilisation and manipulation within the domestic context.
  - *Socio-Cultural Identity Politics:* There has been an increase in anti-feminist and anti-LGBT rhetoric, which is often tied to socio-cultural identity politics.
  - *Geopolitical Solidarity:* Pro-Palestinian content is frequently amplified to align domestic political agendas with broader Islamic solidarity movements.

- Effective mitigation of the impact of organised influence operations could involve the following strategic countermeasures:

  - *Fostering Digital Resilience:* Prioritising media literacy and community-based resilience is essential to immunising the public against manipulative tactics.
  - *Preserving Institutional Integrity:* Stakeholders must resist the impulse to deploy "ethical cyber troopers" or retaliatory manipulation. The adoption of deceptive tactics, regardless of the underlying intent, erodes public trust and undermines democratic norms.
  - *Prioritising Authentic Engagement:* Counter-influence efforts are most effective when deployed on platforms that emphasise community and

authenticity, as these environments are more conducive to genuine discourse than traditional, anonymous forums.

Threats and Challenges of Information Manipulation Facing Taiwan

Summer Chen, *Founder, FactLink*

Taiwan is currently contending with increasingly persistent influence operations orchestrated by the People's Republic of China (PRC). These activities represent a hybrid approach that combines overt military intimidation with sophisticated information warfare and psychological tactics to maximise their strategic impact. Notable instances of this tandem approach include the manoeuvres following the 2022 visit of U.S. House Speaker Nancy Pelosi and the drills conducted in response to President Lai Ching-te's international diplomatic engagements.

- The PLA's operational framework includes a psychological warfare component that specifically targets democratic values and human rights discourse within Taiwan. The efficacy of these narratives often rests on their emotional resonance, which can exert a lingering psychological impact on the populace even after the underlying claims have been empirically debunked.

- Key influence operations methods include:

    - *Media and Influencer Collaboration:* Strategic partnerships with pro-Beijing media outlets and digital influencers help amplify narratives. One prominent example is the persona "Yaya" (Liu Zhenya), a mainland Chinese influencer who has garnered significant attention for advocating the reunification of Taiwan via military intervention.
    - *Synthetic Media and False Flag Operations:* The deployment of AI-generated content and deceptive false flag operations aim to undermine the credibility of pro-Taiwan groups and foster internal societal divisions. The proliferation of AI-driven disinformation further poses a threat in diminishing the effectiveness of traditional fact-checking mechanisms.
    - *Cyber-Enabled Information Leaks:* Hacking operations designed to extract and leak sensitive materials are frequently timed during electoral cycles to damage the reputations of pro-Taiwan candidates, and to influence voter perception.

Taiwanese civil society has responded to these multifaceted threats through several established initiatives, including the implementation of comprehensive fact-checking protocols and community-level media literacy programmes. Taiwan also actively participates in global alliances, such as the International Fact-Checking Network and the UkraineFacts alliance, to share best practices and coordinate international responses to disinformation.

- The following strategic countermeasures could help further strengthen national resilience:

    - *Transparency and Timeliness:* Maintaining public trust necessitates information transparency and countering adversarial narratives
    - *Empathetic Crisis Communication:* Tailored responses can address the underlying psychological anxieties of the public, ensuring communication is both relevant and effective.
    - *Advanced Analytical Strategies:* Moving beyond the simple labelling of fake news, to adopt more nuanced strategies that account for diverse emotional and psychological dimensions of information manipulation.

- *Proactive Resilience Building:* Prioritising pre-bunking initiatives and deep community engagement is vital for building societal resilience.

Key points noted from the Q&A session

Issue: AI and algorithmic dynamics reshape the threat landscape, requiring adaptive strategies.

The proliferation of AI has fundamentally altered the threat landscape, requiring the development of highly adaptive strategic responses. While current methodologies exist to mitigate disinformation campaigns, these remain largely reactionary and often prioritise management of viral content. AI is diminishing the ability to distinguish between coordinated state-sponsored operations, and organic grassroots activity.

Issue: False flag operations are becoming increasingly common in both East Asia and Pacific Island regions.

False flag operations may involve malicious actors impersonating pro-government entities to mislead the public and discredit legitimate institutions. Such tactics exploit systemic vulnerabilities within national governance and media ecosystems. It is crucial that governments scrutinise even seemingly supportive digital content. Cultivating deep collaboration between state agencies, non-governmental organisations and international monitoring networks is essential.

Issue: The risk of short-term countermeasures mirroring disinformation tactics.

Mirroring adversarial disinformation tactics as a countermeasure is flawed, as the adoption of deceptive methods raises significant ethical concerns. It may risk a cycle of escalation which could further destabilise the information environment. Long-term

countermeasures should focus on integrating pre-bunking strategies and media literacy education into school curriculum; strengthening media ecosystems and civil society capacity through funding and training; and encouraging investigative journalism. It should also involve pushing proactive narratives and transparency, while communicating in multiple languages and platforms.

Panel 4: Strategies to Manage Disinformation

The Centre for Humanitarian Dialogue's (HD) Information Threats Diplomacy: Mediating the Digital Frontlines of Conflict

Jacobo Quintanilla, *Programme Manager, Social Media and Conflict Mediation*

Information manipulation has evolved into "cognitive warfare," representing a systemic security threat that has largely outpaced the capabilities of traditional diplomacy. While fact-checking and digital literacy remain essential for addressing the disinformation supply chain, they are increasingly insufficient and are often reactive. These efforts must be complemented by "discrete diplomacy" that fosters online restraints and uses the information arena as an avenue for conflict prevention.

- Information manipulation poses a significant threat to social cohesion and conflict prevention. Technologies originally designed to connect us are now frequently used to divide societies, creating an infrastructure that shapes attention and weakens critical thinking.

- The information space remains largely ungoverned by diplomatic norms, becoming a primary arena for confrontation where online manipulation fuels violence and erodes the trust necessary for dialogue. Threats to security have intensified through AI-powered manipulation, reduced platform moderation, and the spread of content into private messaging apps. Automated influence operations, deepfakes, and influence-for-hire services have now become mainstream tools of statecraft and hybrid warfare.

- Current diplomatic responses remain largely reactive, focusing on regulation, fact-checking, and digital literacy. While these measures are essential for addressing the disinformation supply chain, they are insufficient on their own. A more proactive diplomatic approach is required, specifically one centred on conflict prevention and social cohesion. This strategy must promote online restraint and engage all stakeholders to develop shared principles of responsible behaviour, effectively complementing existing cyber and arms control frameworks.

- There are three preventative strategies against information manipulation:

    - Establishing voluntary "Social Media Codes of Conduct" for political actors,
    - Creating crisis communication channels to de-escalate tensions, and
    - Developing regional norms.

- ASEAN can lead the way in defining rules against cross-border influence operations to protect social cohesion.

- HD's approach to diplomacy in the information environment is pragmatic, grounded in experience, focused on risk reduction, and aimed at preventing

escalation in an increasingly contested information space. For the past five years, HD has been experimenting with discrete, behind-the-scenes diplomacy on technology issues around cyber, AI, social media, and most recently outer space. Accomplishments of HD include the following:

- HD has facilitated social media codes of conduct and brought together political parties to agree on voluntary norms for responsible social media use in places such as Thailand, Indonesia, Kosovo and Bosnia. These context-specific codes help fill gaps where regulation is weak or contested and collectively clarifies what constitutes acceptable behaviour online.
- HD established communication channels to de-escalate disinformation by setting up discrete channels between actors, including those across borders, to enable real-time debunking or cross-checking important information and facilitate information exchange on potential harmful narratives. These channels could be official or unofficial and function like crisis hotlines for cybersecurity incidents, preventing escalation during moments of tension
- HD explored regional norms for responsible behaviour online and opportunities to help develop regional norms for online behaviour within ASEAN. While ASEAN has made important progress through guidance, capacity building, and voluntary principles, including efforts under the ASEAN Defence Ministers' Meeting Cyber and Information Centre of Excellence (ACICE), there are no regional norms that explicitly prohibit cross-border disinformation campaigns.

Information Integrity: Forging a Whole-of-Society Defence Against Disinformation

Angie Drobnic Holan, *Director (International Fact-Checking Network), Poynter Institute*

The fact-checking sector has undergone a significant transformation, migrating from the verification of rudimentary fake news to the mitigation of sophisticated, multi-dimensional threats, including state-sponsored disinformation, transnational financial fraud, and generative artificial intelligence. The International Fact-Checking Network (IFCN) has emerged as a central authority in this field, establishing global standards through its Code of Principles. This framework mandates rigorous adherence to non-partisanship, methodological clarity, and transparency in both funding and organisational structure.

- In the current landscape, disinformation is increasingly viewed as a persistent systemic challenge to be managed through ongoing vigilance, rather than an issue to be solved. The primary concerns currently facing global fact-checking organisations include the following:

    - State-Sponsored disinformation is increasingly deployed as a primary instrument of statecraft, with coordinated campaigns frequently accompanying military conflicts to manipulate geopolitical narratives.

- There is a surge in transnational digital crime, including sophisticated financial scams and fraud proliferating through encrypted messaging and social media, often utilising false digital claims to defraud individuals of their life savings.
- Generative AI is significantly accelerating the cycle of information manipulation, producing synthetic media at a scale and speed that threatens to overwhelm traditional manual verification processes.
- State-sponsored disinformation is increasingly deployed as a primary instrument of statecraft, with coordinated campaigns frequently accompanying military conflicts to manipulate geopolitical narratives.
- There is a surge in transnational digital crime, including sophisticated financial scams and fraud proliferating through encrypted messaging and social media, often utilising false digital claims to defraud individuals of their life savings.
- Generative AI is significantly accelerating the cycle of information manipulation, producing synthetic media at a scale and speed that threatens to overwhelm traditional manual verification processes.

- Factchecking entities worldwide are adopting innovative strategies to counter these evolving threats:

  - VoxCheck in Ukraine uses narrative tracking which organises debunks into broader false narratives (e.g. “Ukraine isn’t a real state”). This can improve media literacy by inoculating the public against specific narrative themes and not just individual fakes.
  - Lupa in Brazil created a specific database/tool called “Is it a scam?” to combat financial fraud. This addressed deepfakes used for fundraising scams (e.g. manipulated video of Bolsonaro).
  - VERA Files in the Philippines partnered with Meta’s third-party fact-checking programme and found that AI is “supercharging” scams. Meta ended this program in the US though it continues globally.
  - Maldita in Spain combines factchecking with a public policy/ advocacy arm and lobbies government on “information integrity” measures, breaking the traditional practice of journalists staying out of policy.
  - Full Fact in UK is developing AI tools that will be able to automate the detection of repeated misinformation across text, video and audio.
  - Global Fact Check Bot is an LLM chatbot that is entirely focused on fact-checking content to see if they could create a model that could confine itself to fact-checked information without hallucinations. It has not been released publicly but their website can be accessed.
  - Newschecker in India is working in 13 languages (most factcheckers work in four or five languages). This organisation is able to look at AI and compare the languages in an effective manner.
  - MediaWise in the United States is an example of a fact-checking organisation that focuses on media literacy. They have teen-oriented videos and AI guidance for newsrooms. They have also done media literacy for librarians.

- Steps that have to be taken to safeguard the future of the information ecosystem include the following:

    - Regulatory frameworks should mandate algorithmic transparency and provide access to platform data for independent researchers.
    - More investment should be directed towards information integrity, supporting high-quality journalism and robust factchecking as essential infrastructure.
    - Given the borderless nature of digital threats, the formation of international coalitions involving governments, civil society, and the private sector is vital for collaborative response.

Restoring Trust in the Age of Disinformation: Lessons from Online Harms Governance

Dr Chew Han Ei, *Head (Governance and Economy), Institute of Policy Studies*

Singapore has adopted an expansive legislative and normative framework to combat online harms, utilising a multifaceted approach that integrates rigorous statutory requirements with broader efforts to cultivate digital trust. The "Sender-Message-Channel-Receiver" (SMCR) model can assist to serve as a conceptual framework to categorise various legal instruments based on their specific intervention points within the information life cycle.

- The restoration of confidence within the digital ecosystem is a critical priority, extending beyond the mitigation of disinformation to the broader establishment of digital trust. This trust is analysed through both mechanical and human dimensions:

    - Mechanical Dimensions: These organisational factors encompass the traditional pillars of safety, security, privacy, and transparency, alongside user experience considerations such as content quality and usability. Emerging priorities in this domain include interoperability, algorithmic fairness, and redressability (defined as the restorative capacity to rectify harm when systemic failures occur).
    - Human Dimensions: These encompass relational traits, community values, and network ties. Digital literacy remains the most decisive factor here. Empirical analysis demonstrates the quadratic relationship of digital literacy where trust increases rapidly as basic literacy improves, though it may diminish among professionals (e.g., IT professionals) due to heightened scepticism. Consequently, interventions targeting lower literacy levels often yield disproportionately high returns in societal trust.

- The long-term management of the digital information environment necessitates a tripartite strategy that involves:

    - the continuous measurement of trust through established metrics,
    - the enactment of proportionate legislation, and

    - comprehensive public education to ensure these legal frameworks are understood and accessible.

- With regard to legislation, the complexity of Singapore's regulatory landscape necessitates clear categorisation to avoid public cognitive overload. The SMCR model can provide this clarity by organising the diverse array of existing and forthcoming regulations:

    - *Sender-Oriented Regulation:* Singapore's Foreign Interference (Countermeasures) Act (FICA) targets perpetrators by empowering authorities to issue directions against hostile actors, a mechanism utilised in 2023 to neutralise coordinated accounts.
    - *Message-Oriented Regulation:* The Criminal Law (Miscellaneous Amendments) Bill provides precision safety by clarifying that obscene content includes synthetic media such as deepfakes, criminalising large-scale distribution while respecting private consensual exchanges.
    - *Channel-Oriented Regulation:* Singapore's Online Criminal Harms Act (OCHA) mandates that platform providers and internet service providers restrict access to criminal content, such as fraudulent scam sites.
    - *Receiver-Oriented Regulation:* Singapore's Online Safety Relief and Accountability Act focuses on victim protection and recourse, establishing a dedicated Online Safety Centre to handle grievances related to egregious harms, including child safety and malicious smears.

Key points noted from the Q&A session

Issue: The necessity of strengthening cross-border cooperation and institutional accountability.

The mitigation of digital threats requires a robust framework for international collaboration that transcends disparate value systems and legal structures. In-person engagement continues to be essential for building high-level trust that is necessary for such cooperation, as purely digital interactions often suffer from inherent scepticism. Success is most likely achieved when nations focus on shared regional interests and minimum common denominators, including egregious harms, such as Child Sexual Abuse Material (CSAM), which even non-aligned parties can reach a consensus. The existing six country collaboration in Southeast Asia regarding CSAM serves as a functional precedent and demonstrates that collective action on universally condemned risks can establish the foundational trust required to eventually address more politically sensitive content.

Issue: The evolution of platform accountability and the enduring relevance of factchecking.

Fact checking remains crucial due to the public's fundamental requirement for verification. Platforms utilise various liability models; for instance, Meta maintains political distance by outsourcing verification to third parties, whereas TikTok internalises moderation, treating factcheckers as merely one of several analytical

signals. Decentralised models such as Community Notes may falter in polarised environments without rigorous transparency. A multi-layered verification strategy (e.g., incorporating professional factcheckers, community notes, and Trusted Flaggers) offers a more resilient approach. However, any rule-based or algorithmic system remains susceptible to the risk of exploitation of sophisticated actors, thus requiring a state of constant operational vigilance.

Issue: Divergent national regulatory frameworks and the complexities of enforcement.

Significant differences exist between global regulatory models, complicating the enforcement of information standards. While the European Union has adopted rigorous frameworks like the Digital Services Act, jurisdictional limitations in other regions often limit action against international bad actors. In authoritarian contexts, global standards like those of the IFCN may be difficult to implement if local laws prohibit the scrutiny of state-backed narratives. Consequently, Public-Private Partnerships (PPPs) and soft power initiatives serve as essential alternatives to traditional regulation, enabling governments to negotiate directly with technology firms for the removal of specific harms when statutory measures prove insufficient to match the scale of the problem.

Issue: Managing the information landscape through continuous adaptation.

As technology evolves, regulatory frameworks risk lagging innovation. Society must adopt a mantra of continuous adaptation, rather than seeking a definitive legislative resolution. Current trends, particularly political shifts towards deregulation and the AI-driven acceleration of disinformation, suggest that the information environment is becoming increasingly volatile. Strategic focus should shift towards identifying diplomatic, regulatory, and technical levers to manage and reduce systemic risk in the digital ecosystem.

Issue: The integration of community-led initiatives as a pillar of information integrity.

A fundamental shift from top-down intervention to bottom-up, community-owned solutions is required to ensure long-term sustainability. Information integrity initiatives are most effective when they respect dissenting voices to avoid pushing them into radicalised echo chambers. Moving away from models that treat users as mere attention suppliers, new community-led frameworks, such as Singapore's community-led Checkmate, which operates on WhatsApp and Telegram, integrating professional fact checkers with citizen participation. By rewarding users for identifying and forwarding dubious content for verification, these models foster a sense of collective ownership and community pride in maintaining regional information hygiene.

About the S. Rajaratnam School of International Studies

The S. Rajaratnam School of International Studies (RSIS) is a global graduate school and think tank focusing on strategic studies and security affairs. Its five Research Centres and three Research Programmes, led by the Office of the Executive Deputy Chairman, and assisted by the Dean on the academic side, drive the School's research, education and networking activities.

The graduate school offers Master of Science Programmes in Strategic Studies, International Relations, International Political Economy and Asian Studies. As a school, RSIS fosters a nurturing environment to develop students into first-class scholars and practitioners.

As a think tank, RSIS conducts policy-relevant and forward-looking research in both national and international security, science and technology, society and economic and environmental sustainability. RSIS also produces academic research on security and international affairs. It publishes scholarly research in top-tier academic journals and leading university presses, and distributes policy research in a timely manner to a wide range of readers.

Centre of Excellence for National Security (CENS)

The Centre of Excellence for National Security (CENS) is a research centre that studies, publishes, and speaks publicly on national security areas including cybersecurity, cyber conflict, disinformation, online harms, hybrid threats, foreign interference, economic interference, social resilience, radicalisation, and the impact of technology (including emerging technologies like artificial intelligence) on them, especially from the perspective of small states and non-western regions such as Singapore and Southeast Asia.